JC07 Rec'd PCT/PTO 21 DEC 2001

FORM PTO-1390 (Modified) (REV 11-2000) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER
11016-0007

U.S APPLICATION NO. (IF KNOWN, SEE 37 CFR
10/018661

INTERNATIONAL APPLICATION NO	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/FR00/01659	**15 JUNE (15.06.00)**	**23 JUNE 1999 (23.06.99)**

TITLE OF INVENTION
METHOD OF EVALUATING THE QUALITY OF AUDIO-VISUAL SEQUENCES

APPLICANT(S) FOR DO/EO/US
Jamal BAINA and Pierre BRETILLON

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371
3. ☐ This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include itens (5), (6), (9) and (24) indicated below.
4. ☒ The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. ☒ A copy of the International Application as filed (35 U.S C. 371 (c) (2))
   - a. ☐ is attached hereto (required only if not communicated by the International Bureau).
   - b. ☒ has been communicated by the International Bureau.
   - c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☒ An English language translation of the International Application as filed (35 U.S C. 371(c)(2)).
   - a. ☒ is attached hereto.
   - b. ☐ has been previously submitted under 35 U.S.C. 154(d)(4).
7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371 (c)(3))
   - a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   - b. ☐ have been communicated by the International Bureau.
   - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   - d. ☐ have not been made and will not be made.
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371 (c)(4)).
10. ☐ An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371 (c)(5)).
11. ☐ A copy of the International Preliminary Examination Report (PCT/IPEA/409).
12. ☐ A copy of the International Search Report (PCT/ISA/210).

Items 13 to 20 below concern document(s) or information included:

13. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
14. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3 28 and 3.31 is included.
15. ☐ A **FIRST** preliminary amendment.
16. ☐ A **SECOND** or **SUBSEQUENT** preliminary amendment.
17. ☐ A substitute specification.
18. ☐ A change of power of attorney and/or address letter.
19. ☐ A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.
20. ☐ A second copy of the published international application under 35 U.S.C. 154(d)(4).
21. ☐ A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).
22. ☐ Certificate of Mailing by Express Mail
23. ☒ Other items or information:

amended English translation under Article 34
corrected pages 24 and 25 with claims 1-6

PCTUS1/REV03

10/018661 JC13 Rec'd PCT/PTO 21 DEC 2001

U.S. APPLICATION NO. (IF KNOWN, SEE 37 CFR	INTERNATIONAL APPLICATION NO	ATTORNEY'S DOCKET NUMBER
10/018661	PCT/FR00/01659	11016-0007

24. The following fees are submitted:.

	CALCULATIONS	PTO USE ONLY
BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)) :		
☐ Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO **$1040.00**		
☒ International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO **$890.00**		
☐ International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO **$740.00**		
☐ International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) **$710.00**		
☐ International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) **$100.00**		
ENTER APPROPRIATE BASIC FEE AMOUNT =	**$890.00**	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☒ 30 months from the earliest claimed priority date (37 CFR 1.492 (e))	**$130.00**	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total claims	6 - 20 =	0	x $18.00	**$0.00**	
Independent claims	1 - 3 =	0	x $84 00	**$0.00**	
Multiple Dependent Claims **(check if applicable)**.			☐	**$0.00**	
TOTAL OF ABOVE CALCULATIONS =				**$1,020.00**	
☐ Applicant claims small entity status. See 37 CFR 1.27). The fees indicated above are reduced by 1/2.				**$0.00**	
SUBTOTAL =				**$1,020.00**	
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492 (f)). +				**$0.00**	
TOTAL NATIONAL FEE =				**$1,020.00**	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31) **(check if applicable).** ☐				**$0.00**	
TOTAL FEES ENCLOSED =				**$1,020.00**	
				Amount to be: refunded	$
				charged	$

a. ☐ A check in the amount of ________ to cover the above fees is enclosed.

b. ☒ Please charge my Deposit Account No. **50-1088** in the amount of **$1,020.00** to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. **50-1088** A duplicate copy of this sheet is enclosed.

d. ☐ Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Christopher W. Brody
Clark & Brody
1750 K Street, NW, Suite 600
Washington, DC 20006

Telephone 202-835-1111
Facsimile: 202-835-1755

Christopher W. Brody

SIGNATURE

Christopher W. Brody

NAME

33,613

REGISTRATION NUMBER

December 21, 2001

DATE

10018661 .121401 018661
JC13 Rec'd PCT/PTO 21 DEC 2001

A METHOD OF EVALUATING THE QUALITY OF AUDIOVISUAL SEQUENCES

The present invention relates to a method of evaluating the quality of an audiovisual sequence, such a sequence being defined in its most general form as comprising audio and/or video signals.

BACKGROUND OF THE INVENTION

The ability to digitize audio and video signals opens up the possibility of being able to copy, store, or transmit this type of information while maintaining constant quality. However, the large quantity of information conveyed by audiovisual signals makes it necessary in practice to use digital compression methods in order to reduce the bit rate.

The Motion Picture Experts Group standard MPEG2 describes techniques of a certain type that are applicable to reducing bit rate. Those algorithms are said to be "with loss", since the signals played back after decoding are no longer identical to the original signals. In order to maintain acceptable quality for the final viewer, algorithms for reducing bit rate take advantage of perceptual properties of human eyes and ears. In spite of this, signal content and the constraints imposed on bit rate or bandwidth available for transmission mean that characteristic degradation appears in the signal after decoding. Such degradation introduced by the global MPEG2 system for encoding and transmission has a direct influence on quality as finally perceived.

Automatic evaluation of the quality of audiovisual signals has a wide range of applications in digital television: production, distribution, and performance evaluation of systems.

Unfortunately, existing apparatuses are designed for laboratory tests and are unsuited for remote surveillance of distribution networks.

There are two different ways of qualifying the degradations that affect picture and sound quality during application of encoding to reduce bit rate or during transmission. Firstly, subjective tests conducted under precise conditions provide results that are reproducible. However such tests are lengthy and expensive to perform. Secondly, automatic systems for evaluating quality on the basis of objective measurements make it possible, for example, to facilitate the development of encoding algorithms and the making of comparisons between them. Such systems make it possible to test digital systems on a spot basis or continuously. In order to obtain objective measurements that are significantly correlated with subjective values, the properties of the human visual system must first be taken into account.

The notion of quality is essentially relative. Even a viewer placed under ordinary conditions of observation (at home) judges the quality of signals made available by comparison with a reference. Under such circumstances, the reference is constituted by the viewer's expectations or habits. Similarly, a method of objectively evaluating quality analyzes degradations introduced by the system on the signals by taking account of reference signals that are present at the input to the system. The study of objective metrics thus requires firstly analysis of defects introduced in the signals, and secondly analysis of the human perceptual system and its properties. The various approaches are based either on computing an error signal, or on identifying signatures that are specific to artifacts introduced by the audiovisual system. The application of perception models makes it possible to evaluate the importance of degradations for the human perceptual system HPS.

Subjective tests are the result of submitting audiovisual signals to a panel of observers representative of the population. A set of satisfaction tests is performed under controlled viewing and listening

conditions. The signals are presented to the observers under a predefined protocol to enable the observers to respond to final quality. Quality is graded using a predefined scale. Quality evaluation scores are obtained after presenting audio, video, or simultaneous audio and video sequences. Statistical computations serve to refine the individual scores by filtering them and homogenizing them. Various subjective test methodologies have been standardized, in particular in the International Telecommunications Union Recommendation ITU-R Bt.500 entitled "Method for the subjective assessment of the quality of television pictures". Two such methodologies using a continuous scoring scale are;

· DSCQS: "double stimulus continuous quality scale" protocol; and

· SSCQE: "single stimulus continuous quality evaluation" protocol.

The first method serves to obtain a score for a 10-second video sequence. Two sequences A and A' are presented in succession corresponding respectively to the original and to the degraded sequence (cf. Figure 1).

The second method omits reference signals and evaluates a given sequence in intrinsic manner. Figure 2 shows a curve of subjective scores achieved during a 30-minute long sequence. The abscissa axis represents time. A subjective score sample was taken every N seconds. The ordinate represents the quality grading scale. The curve shows the impact on subjective quality of all of the disturbances to which the sequence was subject.

Objective measurements can be performed using various approaches.

The principle of the approach which uses perception models is to stimulate the behavior of the human perceptual system (HPS) in part or in full. Given that in this context it is the quality of audiovisual signals that is to be determined, it suffices to evaluate the perceivability of errors. By modeling certain functions

of the HPS, it is possible to quantify the impact of errors on the sense organs of humans. These models act like weighting functions applied to the error signals. In this way, the effect of each degradation is modulated in proportion. The overall process makes it possible objectively to evaluate the quality of signals passing through an audiovisual system (see Figure 3).

Reference signals S_{ref}, e.g. representing an audiovisual sequence, and signals S_0 from said sequence and degraded by an audiovisual system SA are compared in a module MID for identifying defects, and then a score NT is given to the defects by comparison with a model MOD.

In the context of computing an error signal, signal-to-noise ratio can be considered as a quality factor. However it is found in practice that it is poorly representative of subjective quality. This parameter is very general, and thus incapable of spotting local degradations, of the kind that are typical in digital systems. Furthermore, signal-to-noise ratio makes it possible to evaluate the fidelity of degraded signals compared with the original very strictly, but that is different from evaluating the overall perceived quality.

To obtain a better evaluation of quality, it is necessary to use a large amount of experimental data concerning the human perceptual system. Application of the data is greatly facilitated since the system has been studied in terms of its sensitivity to a stimulus (in this case the error) in the context of a picture, for example. In this context, what matters is the response of the human visual system (HVS) to a contrast and not to an absolute magnitude such as luminance.

Various test images, such as uniform areas of luminance, or frequencies in space or time, have made it possible experimentally to determine the sensitivity of the visual system and the associated values of just-perceivable contrast. The appearance of the HVS response to light intensity is logarithmic, with optimum

sensitivity being at spatial frequencies close to 5 cycles/degree. Nevertheless, those results need to be applied with prudence, since they are visibility threshold values. This explains why it is difficult to predict the importance of degradations of large amplitude.

Hearing models proceed in similar manner. Experimentally, the sensitivity to various stimuli is measured. It is then applied to various signal errors in order to evaluate quality.

However, audiovisual signals are complex in terms of richness of information. Furthermore, in practice, the use of that type of model to evaluate audiovisual signals raises several problems. In addition to the fact that the reference and degraded signals need to be available physically at the same location, it is also essential for sequences to be caused to correspond exactly in space and in time. That approach is therefore applicable to evaluating equipment such as an encoder when all of the equipment is located in a single laboratory, or to some cases of evaluating transmission such as satellite transmission where the transmitter and the receiver can both be on the same premises.

The approach which makes use of parametric models combines a series of parameters or degradation indicators chosen for generating an overall objective score.

The objective measurements applied to the audio and/or video signals are indicators of signal content and of the degradations to which they have been subjected. The relevance of these parameters depends on how representative they are in terms of sensitivity to defects.

Two categories of approach are then possible when generating parameters:

1) category I: "with a priori knowledge of the reference signal"; and

2) category II: "without a priori knowledge of the reference signal".

The first approach category I relies on performing the same transformation or the same parameter computation on the reference signal and on the degraded signal. Generating an overall quality score relies on comparing results coming from both treatments. The measured difference represents the degradation to which the signal has been subjected.

The second approach category II does not require knowledge about the original signal, but only knowledge about the characteristics which are specific to degradation. It is then possible to compute one indicator for one or more degradation types. Low bit rate encoding and disturbed broadcasting of digital television signals generate identifiable characteristic defects: the blocking effect, picture freezing, etc. Factors for detecting these defects can be generated and used as quality indicators.

An example of a parametric model:

Numerous parameters have been proposed in the literature for implementing parametric models. The present invention does not seek to define new parameters, but to propose a general model for making use of such measurements.

The approach consists in comparing two images (the reference image and the degraded image) only on the basis of parameters that are characteristic of their content. Which parameters are selected is associated with their sensitivity to certain degradations which the system under evaluation produces. Thereafter, a quality measurement is built up by correlation using a series of objective measurements.

As an example, we mention a technique developed by the US Institute of Telecommunication Sciences (ITS). It relies on extracting a space parameter SI and a time parameter TI characteristic of sequence content (see

Figure 4). For further information, reference can be made to an article by A.A. Webster et al. entitled "An objective video quality assessment system based on human perception", published in SPIE, Vol. 1913, pp. 15-26, June 1993.

The space information that is considered as being important in this case is outline information. For an image I at time t, the space parameter SI is obtained from the standard deviation of the image as filtered by Sobel gradients. This technique reveals the outlines of the image under analysis, and these play an important part in vision:

$$SI_t = \sigma_{x,y}(\mathrm{Sobel}[I_t(x,y)])$$

In analogous manner, time information at a given instant is defined by the standard deviation of the difference between two consecutive images:

$$TI_t = \sigma_{x,y}(I_t(x,y) - I_{t-1}(x,y))$$

A measurement based on those two pieces of information makes it possible to evaluate change in content between the input of a video system (S_{ref}) and its output (S_s), by using various comparisons.

$$M_1 = \log_{10}\left[\frac{TI_s(t)}{TI_{ref}(t)}\right]$$

$$M_2 = \left[\frac{ST_{ref}(t) - ST_s(t)}{SI_{ref}(t)}\right]$$

$$M_3 = [TI_s(t) - TI_{ref}(t)]$$

The three parameters M_1, M_2, and M_3 are taken from these comparisons by a comparator COMP. Each of them is sensitive to one or more degradations. Thus, by comparing parameters SI, loss of focus (reduction in SI) is taken into account as are the outlines that are artificially introduced by the blocking effect (increase in SI). Similarly, differences between two versions of TI reveal defects in the encoding of motion.

The following step consists in summing M_1, M_2, M_3 over time using one of the Minkowski norms Lp (in general, p=1, 2, or ∞). In this manner, it is possible to construct a summing model. That makes it possible to produce a quality score at the outlet from a summing module SMOD. The chosen model is a linear combination of the terms in M_i:

$$Q = \alpha + \beta M_1 + \gamma M_2 + \mu M_3$$

The weighting coefficients (α, β, γ, μ) are computed by an iterative procedure MIN for minimizing distortion between the objectives scores Q and the subjective scores obtained on the same batch of pictures. The idea is to use iteration to discover the parameters of the combinatory model. In this manner, the estimated objective measurement will come as close to possible to the subjective score. The performance index of the model is given by the correlation coefficient.

An example of a model has been proposed in the literature. It makes it possible to obtain a good correlation coefficient: 0.92.

$$Q = 4.77 - 0.992M_1 - 0.272M_2 - 0.356M_3$$

Nevertheless, it appears that combinatory models produce performance that is not so good when they are used with pictures other than those appearing in the batch which was used for devising the model.

This approach is less constricting to implement than the preceding approach. Nevertheless, in practice, it remains difficult to achieve space and time correspondence between the scores of the two signal sequences.

OBJECTS AND SUMMARY OF THE INVENTION

An object of the invention is to provide a method which makes it possible to achieve good correspondence between objective measurements and subjective scores given by a panel of spectators.

Another object of the invention is to provide a method enabling an audiovisual sequence to be evaluated

in absolute manner, i.e. without having access to a non-degraded original sequence.

Another object of the invention is to provide a method making it simple and effective to evaluate the quality of audiovisual signals in a network for broadcasting audio and/or video signals.

At least one of the above-specified objects is achieved by a method of evaluating the quality of an audiovisual sequence, the method implementing:

a) training, comprising allocating a subjective score NS_i to each of N_0 training sequences S_i (where i = 1, 2, ..., N_0) presenting degradations identified by a training vector MO_i which is given to each sequence S_i in application of a first vectorizing method, in order to build up a database of N_0 training vectors MO_i with subjective scores NS_i;

b) classifying the N_0 training vectors MO_i into <u>k</u> classes of scores as a function of the subjective scores NS_i that have been allocated to them, so as to form <u>k</u> training sets EA_j (where j = 1, 2, ..., k) which have <u>k</u> significant training scores NSR_j allocated thereto;

c) for each audiovisual sequence to be evaluated, generating a vector MO using said first vectorization method; and

d) allocating to the audiovisual sequence for evaluation the significant training score NSR_j that corresponds to the closest training set EA_j.

By way of example, this allocation can be performed by allocating to the sequence, the score which corresponds to the closest training vector.

Preferably, between step b) and c), the following step is performed:

b1) for each training set EA_j, using a second vectorization method to generate a reference dictionary D_j made up of N_j reference vectors VR_l (where l = 1, 2, ..., N_j);

and between steps c) and d):

c1) selecting amongst the reference vectors VR_1 of the k reference dictionaries, the reference vector VR_e which is closest to said vector MO.

In which case, allocation is performed on the basis of the significant training score NSR_j corresponding to the reference dictionary to which said closest reference vector VR_e belongs.

The significant training scores NSR_j can be distributed uniformly along a scoring scale, or better still they can be distributed in non-uniform manner, thus making it possible to make measurements more meaningful, e.g. because at least some of the reference dictionaries can then contain substantially the same numbers of reference vectors.

In a preferred implementation, the significant training scores NSR_j are distributed by a training technique.

The method then comprises, between steps a) and b), identifying k significant training scores NSR_j from subjective scores NS_i each considered as a one-dimensional vector, by finding the minimum distance between the set of the N_0 subjective scores NS_i and the k significant training scores.

BRIEF DESCRIPTION OF THE DRAWINGS

Other characteristics and advantages of the invention appear better from the following description and the drawings accompanying it, in which:

- Figure 1 and Figure 2 show two prior art evaluation methods known respectively as DSCQS and SSCQE;
- Figure 3 shows a known approach implementing perception models;
- Figure 4 shows a method developed by ITS;
- Figure 5 shows a preferred implementation of the invention;
- Figure 6 shows how a dictionary D_j is formed in accordance with the invention;

· Figure 7 shows the process for storing a sequence to be evaluated, in accordance with the invention;

· Figure 8 shows a uniform distribution of subjective score intervals;

· Figure 9 is an example showing the frequency with which subjective scores NS_i appear;

· Figure 10 shows a process for identifying representative subjective scores in accordance with the invention;

· Figure 11 shows an example of a non-uniform distribution of subjective scores as obtained using the process shown in Figure 10;

· Figures 12 and 13 show the precision of evaluation in accordance with the invention; and

· Figures 14 and 15 show an implementation of the invention relating respectively to the training stage and to the operational stage.

MORE DETAILED DESCRIPTION

The invention is presented as an objective method of evaluating the quality of audio and video signals based on a set of parameters.

The method does not require new parameters to be defined. It is based on the idea of proposing a general model based on vector quantification to make use of the measurements. It is a training approach making it possible to give objective quality scores to audiovisual signals. This evaluation is performed in a manner which is correlated with subjective scoring with the help of an objective perception model. To do this, the model operates in two stages.

The first stage is performed on a set of training audiovisual sequences. The model establishes correspondence between two databases derived from the same audiovisual signals:

· a subjective score database (BDSN); and

· a database of objective measurements taken from the audio and video signals (BDMO);

in order to associate a subjective score with a set of degradation indicators (forming a vector of objective measurements).

This stage derives knowledge that is pertinent in qualifying signal quality.

During the second stage, which corresponds to operational use of the method, the method makes use of its knowledge. Each time it is necessary to qualify the quality of an audiovisual sequence, the model extracts parameters representative of degradations. Thereafter, it applies the result of its computations to its knowledge base. This operation makes it possible to give an objective score that is very close to the subjective score which a representative panel would have given. The process used in the present invention makes use of vector quantification. The principle is to search through dictionaries of vectors to find the representative vector that is closest to the parameter vector computed on the audio and video signals. By way of example, the subjective score that is generated can be the score which is associated with the dictionary containing the closest representative vector.

The problems associated with vector quantification have been identified in the literature. They can be summarized by defining its three main interdependent components:

- forming vectors from the information to be encoded;
- forming the dictionary from a training set; and
- determining the closest neighbor using an appropriate distance.

The notion of distance or distortion between two vectors is introduced for finding the closest neighbor in the dictionary. Several distances have been proposed for optimizing vector quantification and for achieving maximum fidelity with the initial signals.

The distance or distortion known as quadratic error is one of the most widely used for vector quantification. The term "distance" is not exact in this case, since mathematically speaking the term constitutes the square of a distance.

$$D(A,B) = \sum_{j=1}^{t}{}' (A_j - B_j)^2$$

where (A, B) are two vectors of dimension t.

Vector quantification is used in the present invention for generating an objective perception model. The model is then used for quantifying the quality of audiovisual signals.

Let there be a set E of N_0 audio sequences S_i of n seconds each. They are all made up of a series of video pictures and of audio samples.

$$E = \{S_i / i = \ldots N_0\}$$

These sequences pass through configurations that are representative of digital television distribution systems. The broadcast and distribution networks used are by satellite, by cable, and by terrestrial network. Disturbances arise during transmission of audiovisual signals and degrade them.

We have performed subjective tests on the above set of degraded sequences. A database of subjective scores has been built.

$$BDES = \{NS_i / i=1 \ldots N_0\}$$

Where NS_i represents the subjective score obtained by sequence S_i in set E.

Furthermore, we have built another database on the basis of objective measurements MO_i performed on the set of sequences E.

$$BDMO = \{LO_i / i=1 \ldots N_0\}$$

with:

$$MO_i = (V_i, \ldots, V_t)$$

Each sequence S_i corresponds to a vector MO_i (see Figure 5). These vectors comprise t parameters V_j

computed from the audio and/or video signals. These parameters can be comparative (category I) or intrinsic (category II). They provide information concerning content and the degradations to which the sequence has been subjected.

In order to form the representative vector MO_i for each audiovisual sequence S_i, a distinct method computes objective parameters that are extracted from samples of the digital audio and video signals.

Using the data as described above, the method begins with a training stage. Appropriate processing of the data serves to develop a knowledge base which is subsequently used by the model during the operational stage.

The set E of sequences S_i is subdivided into k classes of scores EA_j. To do this, use is made of the value of the subjective score NS_i allocated to the sequence S_i. The variation interval of NS_i is thus subdivided into k distinct segments I_j which are associated with the k training sets EA_j. A representative subjective score NSR_j is associated with each segment j. This operation groups data concerning sequences of quality that is judged to be similar or equivalent into each class having a score EA_j.

The value k (e.g. k=5) is taken as an example application in Figure 5. It is possible to envisage distribution over a number of classes that is larger or smaller as a function of the accuracy requirements of the measuring equipment.

The vectors MO_i of objective measurements of the sequences S_i corresponding to a range I_j of subjective score values NS_i are collected together in training set EA_j. k training sets are then formed on the basis of the initial data (cf. Figure 5).

Starting from a training set having M vectors, the reference dictionary which is made up of N vectors is the dictionary which best represents the initial set of

vectors. It uses a group of vectors presenting the smallest mean distance or distortion relative to all of the M vectors in the training set, compared with the other possible candidate dictionaries. Dictionary building is based on forming the best representative vectors.

Classification algorithms are used so as to build a dictionary of representative vectors from an initial set. This is referred to as the "training set".

Several authors have proposed solutions for classifying in dictionaries:

- dynamic swarms or LBG algorithm;
- Kohonen's neural network method.

For each class of scores EA_j and starting from objective measurement vector MO_i together with their scores NS_i (see Figure 6), a procedure FORM is applied to build a dictionary D_j.

<u>k</u> dictionaries D_j, each made up of N_j vectors are associated with the <u>k</u> classes or ranges of subjective scores. The value of N_j is selected as a function of the initial number of vectors in score class EA_j and depending on the accuracy required for the model. Each dictionary D_j is thus associated with an interval I_j of subjective scores.

The algorithms used for building dictionaries D_j are LBG and Kohonen's neural networks. Those methods give comparable results. Those techniques are particularly effective even when the sizes N_j of the dictionaries are deliberately chosen to be small (e.g. N_j = ...), and the reference dictionaries remain representative.

The purpose of automatic apparatus for evaluating signal quality is to provide a final score evaluating said signal. During operational use, the method described by the present invention is performed using two processes (see Figure 7).

The first process lies in treating the audio and/or video samples of the audiovisual sequence to evaluate SAE

so as to extract the parameters therefrom. A vector V_i of audio and/or video quality indicators is formed using categories I and/or II as described above. This serves to present characteristics that are pertinent for qualifying the signals.

The second process (QUANT) uses vector qualification on the vector V_i of input parameters allocated to an audiovisual sequence for evaluation in order to establish a correspondence with the index j of the closest dictionary. For this purpose, the distortion between the incident vector and all of the vectors of the k dictionaries is minimized. That serves to identify the dictionary D_j containing the vector U which is closest to the vector V_i, and thus to identify the index j.

The operation that it is advantageous to use in this approach is vector quantification. It makes it possible to find the closest neighbors of a vector V_i and consequently its best representative in a dictionary or a set of dictionaries. For a given input vector V_i, vector quantification determines which vector of which dictionary is the closest thereto, and allocates the significant training score NSR_j of said dictionary D_j thereto.

The index j is no more than the quality class obtained by grading subjective tests performed on audiovisual sequences. With this technique of subdivision into a plurality of training sets, two important points need to be studied:

· the size of each dictionary; and

· the positions of the score ranges for each dictionary.

The size of each dictionary is of considerable importance. The number of vectors has a direct influence on the representativeness of the dictionary, and consequently on the effectiveness of the vector quantification.

The positions of the score ranges are just as important. It is necessary to know which scores are going to be associated with one another. For example, a large range of scores can be set aside for bad quality, thus ensuring that once quality has dropped to a minimum, quantification will detect it. It is also possible, on the contrary, to retain a small range for bad quality, so that the quantifier will detect bad video quality only when it is strongly degraded.

It can thus be seen that using these two parameters it is possible to influence the way vector quantification occurred. It is also possible to influence this quantification by adding preprocessing on the objective parameters computed from the audio and/or video signals.

The operation of the method is described above as comprising three main steps: firstly objective measurements MO_i are made, then dictionaries D_j are built, and finally a search is made for the dictionary containing the vector which is closest to an objective measurement vector. The model can then allocate the representative subjective score NSR_j associated with the dictionary D_j to the sequence S_i as represented by the objective measurements MO_i by making use of its knowledge base. However, no method of selecting ranges in the subjective score scale has been defined, nor has any method been defined for selecting the representative score NSR_j that is associated with each dictionary D_j. The way in which the subjective score scale is partitioned is an important step since it defines the scores which the model is capable of providing during the operational stage.

As defined above, each class is defined by the training set EA_j of objective measurements, and an interval I_j on the scale of subjective scores NS_i.

When performing subjective tests using a discrete scoring scale, the number of representative scores and of corresponding ranges is naturally limited by the number

of levels that can be given to the score (in general five levels).

For subjective tests using a continuous scoring scale, the possibilities are much more varied: the number of training sets can be arbitrary. Two approaches are then possible: either subjective score intervals I_j are selected arbitrarily, or else an automatic procedure is applied enabling intervals I_j to be selected.

Arbitrary partitioning

An arbitrary selection of intervals for subjective scores NS_i (see Figure 8 for a uniform distribution) has the advantage of requiring no particular resource during the hardware installation of the invention in equipment. Unfortunately, such partitioning which takes no account of the way subjective scores for the sequences of the set E (Figure 9) are distributed in fact runs the risk of defining certain intervals which contain few or no subjective scores NS_i, and a single interval might contain nearly all of the scores.

Such a non-uniform distribution of subjective scores amongst the intervals has two drawbacks for the model:

1) Firstly, whatever the size of the dictionaries and the sensitivity of the parameters $V_{1...t}$ to degradation, the difference between the predicted subjective score and the real subjective score cannot be minimized. The operational stage associates each vector $V_{i...t}$ of objective parameters with the score NRS_p of the closest dictionary D_p. Since the subjective score interval represented by NRS_p has a certain interval width, the mean difference cannot drop below a certain threshold which is a function of the width of the interval. When the corresponding training set EA_p contains the major fraction of the sequences S_i, the model will make very frequent use of the score NSR_p and will thus frequently commit a naming error. The mean performance of the model for this class <u>p</u> of scores will thus be limited by said

interval width, and will be improved by reducing the interval. Consequently, for the class p corresponding to the major fraction of the sequences S_i, it is the mean performance of the model which is limited.

It can thus be seen that it is advantageous to partition using smaller intervals in zones that are dense in terms of the number of subjective scores obtained in the database DBNS.

2) Secondly, an arbitrary approach to positioning can lead to having a non-optimal overall number of vectors for the dictionaries. With this type of partitioning, the training sets EA_j that are formed can be of very different sizes. As a result, for a training set EA_p of large size, the algorithm for the classification stage will need numerous vectors in the dictionary D_p in order to represent EA_p with desired distortion. That is because of the wide diversity of the data to be represented. Partitioning that guarantees that significant unbalance will not be obtained concerning the sizes of the training sets can resolve this point. Furthermore, it is not certain that the smaller sizes of the other training sets will enable the sizes of their dictionaries to be reduced. Overall this leads to an increase in the hardware cost of implementing the method, and to a decrease in the accuracy of the model.

A partial response to this drawback is to perform partitioning empirically, each time a set E of sequences is studied. For this purpose, the scale of scores will be partitioned more finely in those locations where the number of scores NS_i is large.

Nevertheless, it is much more advantageous to apply a procedure that is automatic, and that also enables partitioning to be performed optimally by implementing automatic partitioning that adapts to the statistical distribution of the subjective scores allocated to the set E of sequences S_i.

Arbitrary partitioning is not a priori well adapted to the way in which the subjective scores NS_i are distributed along the subjective score scale. Although the set E of training sequences is representative of degradations, it is generally found that the distribution of the values of NS_i is far from being uniform, e.g. when applied to digital television. Figure 9 shows the frequency with which subjective scores NS_i occur: it can be seen that many scores are close to a high level of quality. High quality classes will thus represent the vast majority of scores whereas the lowest quality class will be nearly empty. Using an automatic procedure for optimum partitioning guarantees that this set DBNS of subjective notes is distributed more fairly, thereby making it possible to obtain better performance from the final model.

This problem is advantageously solved by a method comprising two steps: firstly identifying k representative subjective scores NSR_j, and then selecting the subjective score NSR_j that best represents the subjective score NS_i.

1) k representative subjective scores NSR_j are identified on the basis of the subjective scores NS_i (Figure 10). The method assumes that each score NS_i is a one-dimensional vector, so as to apply a reference dictionary building process. One of the LBG, dynamic swarm, or Kohonen's neural network methods is used in order to obtain the desired number k of representatives NRS_j.

This type of method tends to seek out minimum distortion in the sense of minimum distance Δ between the set of NS_i and the set of NSR_j. It thus responds perfectly to the drawbacks of arbitrary positioning.

2) The training set DBMO is classified into k sets EA_j. For this purpose, data pairs (MO_i, NS_i) are considered, each pair corresponding to a sequence S_i. For each pair, the closest representative subjective score

NSR_j is found in NS_i by applying the vector quantification procedure, thus enabling the index j to be determined. The objective data vector MO_i is then added to the training set EA_j. The creation of sets EA_j amongst which the vectors MO_i are distributed terminates once all of the pairs (MO_i, NS_i) have been processed.

An example of optimal partitioning of the subjective score scale is given in Figure 11 which illustrates the difference compared with Figure 8.

The model is used here to illustrate its possibilities on a digital television program containing degradations. The subjective scores were obtained using the SSCQE protocol, i.e. one score every half-second. It is then considered that the program is made up of a series of as many half-second short sequences S_i as there are scores.

Figure 12 shows together how the subjective score NS and the allocated objective score NSR vary over half an hour. It can be seen that the allocated objective score NSR tracks the subjective score NS very closely (dashed line).

Figure 13 summarizes correspondence between the score predicted by the model and the real subjective score for the same experiment, together with the accuracy of the model. Seven classes are distinguished, corresponding to as many predicted score values (objective score NS on the abscissa, subjective score NSR on the ordinate).

For each class, the graph shows the mean of the real subjective scores (Mean) given by the observers. Good linearity can be seen in the correspondence between the two scores, thus constituting a first performance criterion.

The mean of the real subjective scores (Mean) is also shown lying between two other scores (SDev). For each class these scores give the amplitude relative to the mean of the standard deviation of the subjective

scores corresponding to the objective score of the class. A small value means that the model is accurate. The values obtained for standard deviation are comparable to the performance of the subjective tests which constitute the reference for the model, which is entirely satisfactory.

An implementation of the invention is described below with reference to Figures 14 and 15.

In order to evaluate the quality of audiovisual signals, the method thus implements two stages: a training stage (Figure 14); and an operational stage (Figure 15).

The training stage is performed only once. It consists in building k dictionaries D_j of objective measurements, together with the associated representative subjective scores NSR_j. This stage is implemented:

· firstly from the objective measurement database (BDMO) obtained from audio and/or video signals and a processor (not shown) for computing parameters (MO, objective measurements); and

· secondly a subjective score database (DBNS) obtained from the same audio and/or video signals as the database DBM0 and a set of observers.

The training stage comprises three steps:

1) A dictionary building processor finds the k subjective scores NSR_j representative of the database BDNS.

2) Each vector of the database BDMO is added to one of the k training sets EA_j as a function of the class j to which the score NS of the database BDNS corresponding to the vector belongs. The class j is obtained using a vector-quantification processor which finds the score NSR_j that is closest to the score NS.

3) Finally, each dictionary D_j (Dic 1, ..., Dic k) made up of N_j vectors is obtained from the corresponding training set EA_j using a dictionary-building processor.

Thereafter, the operational stage is applied each time the quality of an audiovisual sequence needs to be predicted. This stage makes use of the knowledge acquired by the model during the training stage. For a vector of objective parameters MO coming from an audiovisual sequence, an objective quality score is computed. The objective parameters MO are supplied by a parameter-computing processor which can be of any kind.

This operational stage can thus comprise two steps:

1) A vector-quantification processor finds the vector U which is closest to the input vector of objective parameters MO from amongst all of the vectors of the dictionaries Dj (Dic 1, ..., Dic k) obtained during the training stage. The processor then provides the number j of the corresponding dictionary.

2) The following step can then allocate the quality score of value NSR_j to the audiovisual sequence.

CLAIMS

1/ A method of evaluating an audiovisual sequence, the method being characterized in that it implements:

a) training, comprising allocating a subjective score NS_i to each of N_0 training sequences S_i (where $i = 1, 2, \ldots, N_0$) presenting degradations identified by a training vector MO_i which is given to each sequence S_i in application of a first vectorizing method, in order to build up a database of N_0 training vectors MO_i including only said identified degradations and subjective scores NS_i;

b) classifying the N_0 training vectors MO_i into k classes of scores as a function of the subjective scores NS_i that have been allocated to them, so as to form k training sets EA_j (where $j = 1, 2, \ldots, k$) which have k significant training scores NSR_j allocated thereto;

c) for each audiovisual sequence to be evaluated, generating a vector MO using said first vectorization method; and

d) allocating to the audiovisual sequence for evaluation the significant training score NSR_j that corresponds to the training set Ea_j containing the vector that is closest to the vector MO in the sense of vector quantification.

2/ A method according to claim 1, characterized in that it comprises:

between steps b) and c):

b1) for each training set EA_j, using a second vectorization method to generate by vector quantification a reference dictionary D_j made up of N_j reference vectors VR_l (where $l = 1, 2, \ldots, N_j$);

and between steps c) and d):

c1) selecting amongst the reference vectors VR_l of the k reference dictionaries, the reference vector VR_e which is closest to said vector MO; and

This translation of an amended page covers the amendments made in the original. However, the page breaks match the translation, so that this page is also a replacement page that fits in with the remainder of the translation.

in that step d) allocates to the audiovisual sequence for evaluation the significant training score NSR_j corresponding to the reference dictionary containing said closest reference vector Vr_i.

3/ A method according to claim 1 or claim 2, characterized in that the significant training scores NSR_j are distributed in uniform manner along the score scale.

4/ A method according to claim 1, characterized in that the significant training scores NSR_j of at least some of the k reference dictionaries are distributed in non-uniform manner along the score scale.

5/ A method according to claim 4, characterized in that said distribution is such that at least some of the reference dictionaries contain substantially the same numbers of reference vectors.

6/ A method according to claim 4 or claim 5, characterized in that it comprises, between step a) and b), identifying k significant training scores NSR_j from subjective scores NS_i each considered as a one-dimensional vector, by finding the minimum distance between the set of the N_0 subjective scores NS_i and the k significant training scores.

CORRECTED SHEET

This translation of an amended page covers the amendments made in the original. However, the page breaks match the translation, so that this page is also a replacement page that fits in with the remainder of the translation.

ABSTRACT

The invention provides a method of evaluating the quality of an audiovisual sequence by:

a) training, comprising allocating a subjective score NS_i to each of N_0 training sequences S_i (where i = 1, 2, ..., N_0) presenting degradations identified by a training vector MO_i which is given to each sequence S_i in application of a first vectorizing method, in order to build up a database of N_0 training vectors MO_i with subjective scores NS_i;

b) classifying the N_0 training vectors MO_i into k classes of scores as a function of the subjective scores NS_i that have been allocated to them, so as to form k training sets EA_j (where j = 1, 2, ..., k) which have k significant training scores NSR_j allocated thereto;

c) for each audiovisual sequence to be evaluated, generating a vector MO using said first vectorization method; and

d) allocating to the audiovisual sequence for evaluation the significant training score NSR_j that corresponds to the closest training set EA_j.

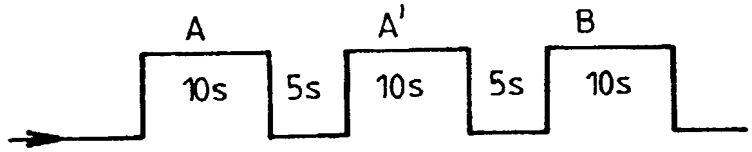

FIG.1

FIG.2

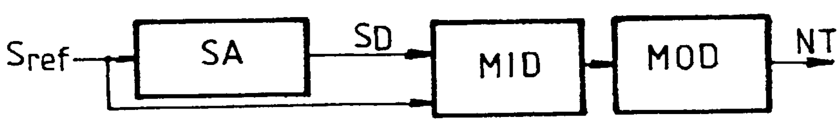

FIG.3

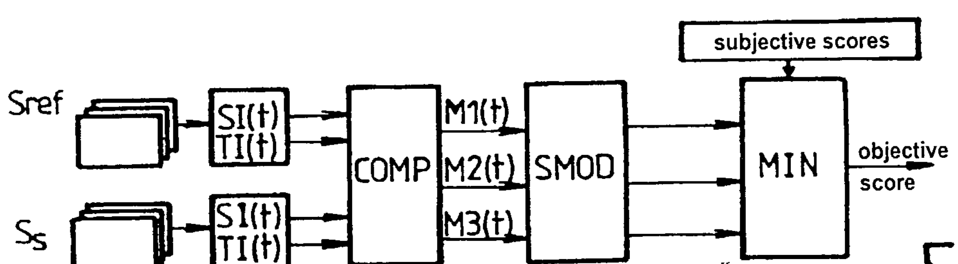

FIG.4

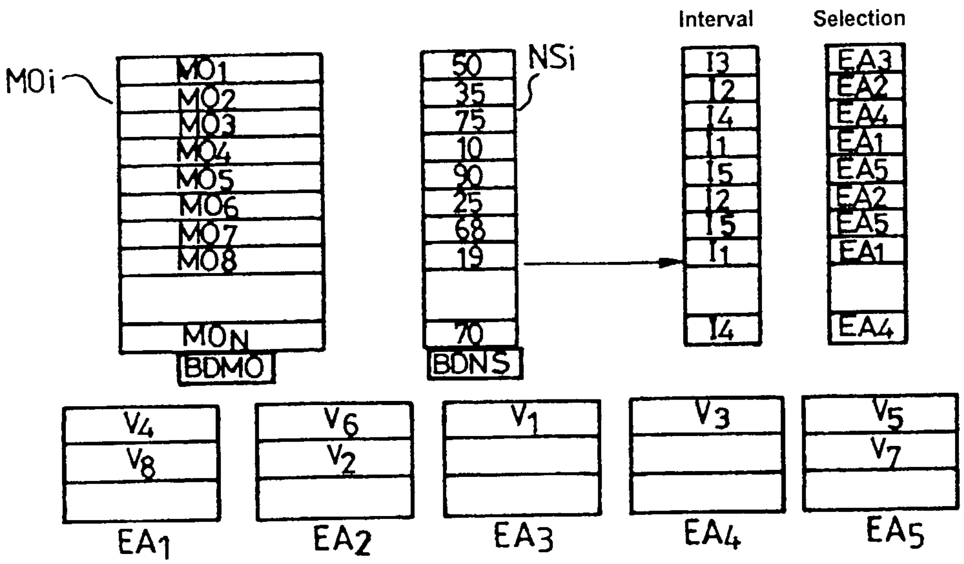

FIG.5

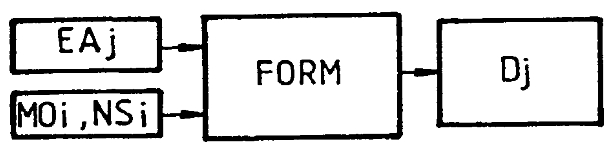

FIG.6

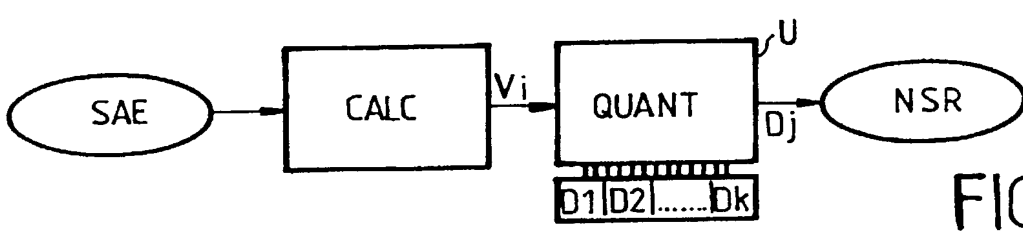

FIG.7

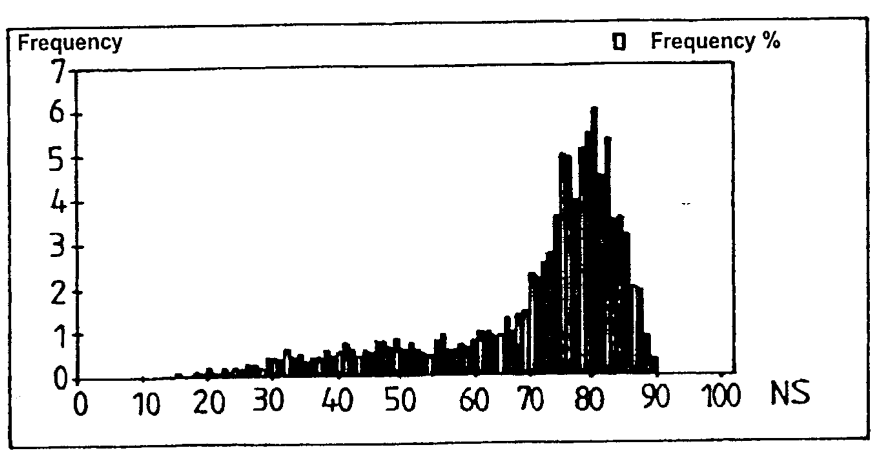

FIG.9

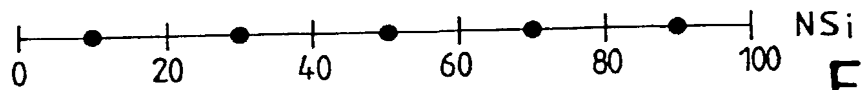

FIG.8

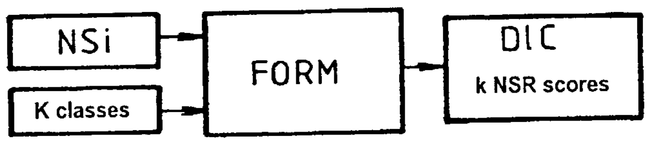

FIG.10

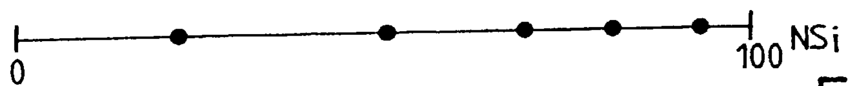

FIG.11

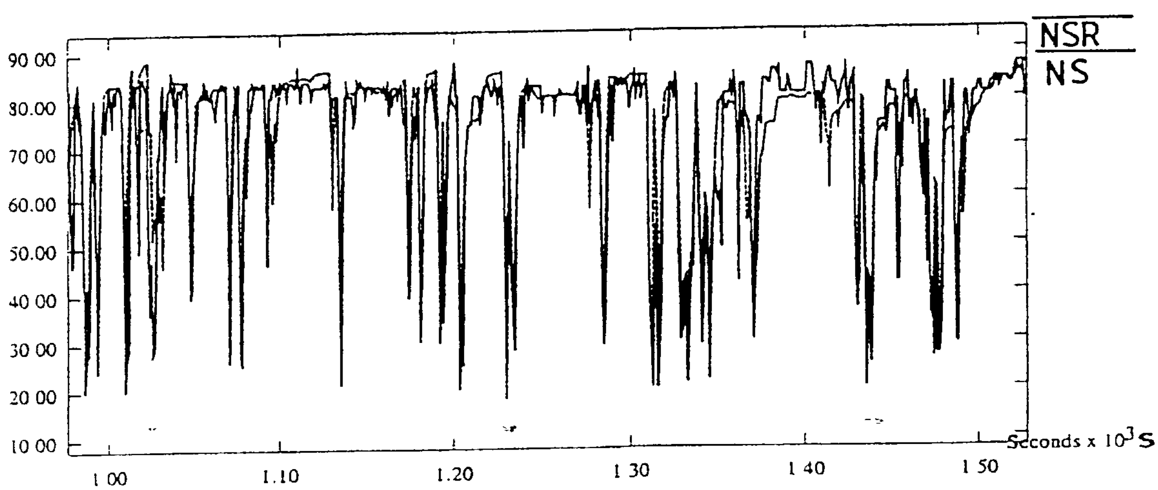

FIG.12

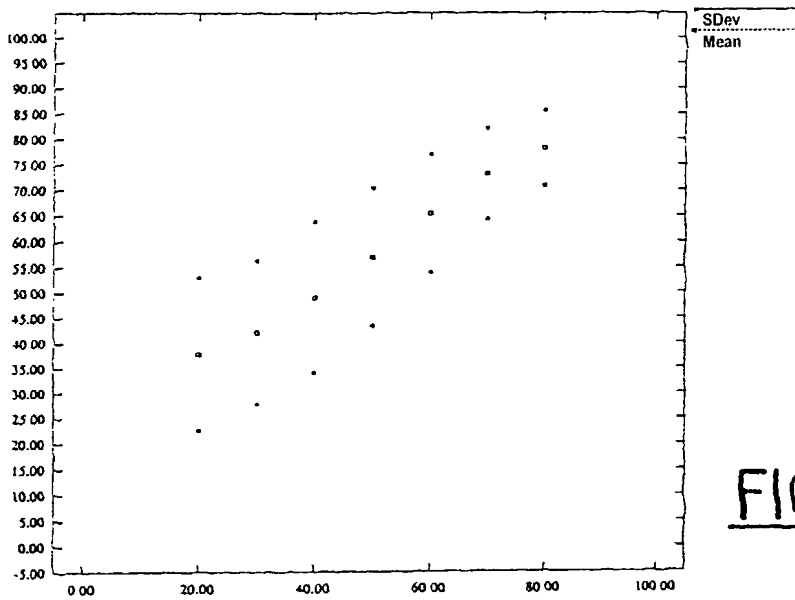

FIG.13

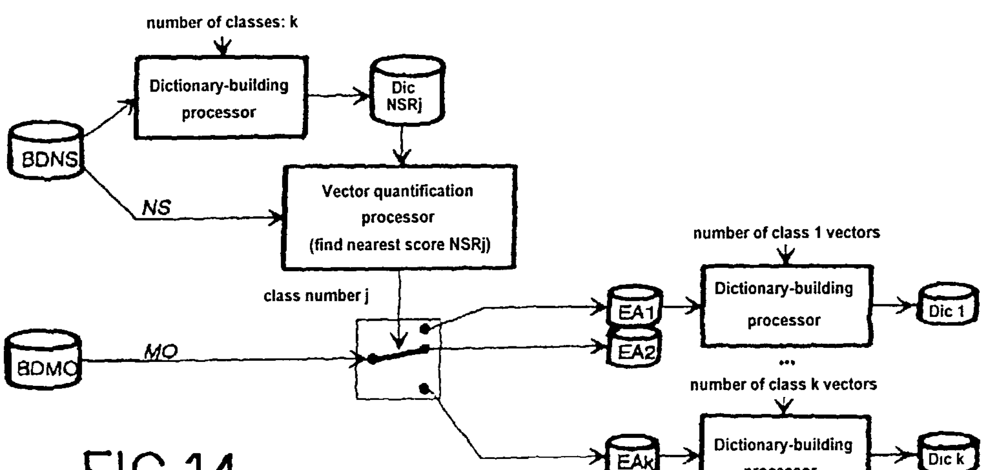

FIG.14

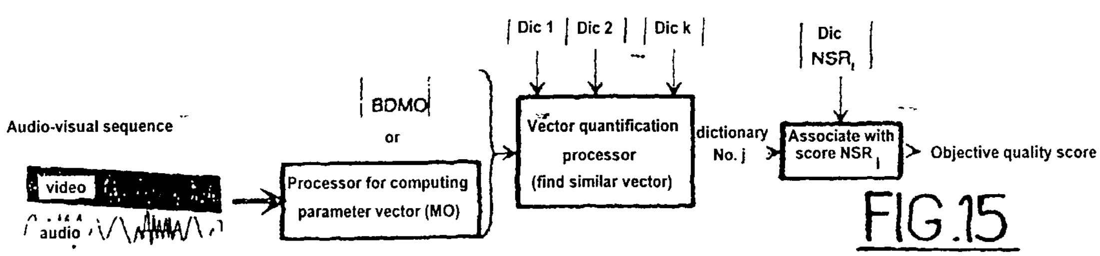

FIG.15

Docket No: 11016-0007

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I declare that:

My residence, post office address, and citizenship are as stated below next to my name. I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first, and joint inventor (if plural names are listed below) of the subject matter that is claimed and for which a patent is sought on the invention entitled

METHOD FOR EVALUATING THE QUALITY OF AUDIO-VISUAL SEQUENCES

☐ the specification of which is attached hereto.
☒ was filed on June 15 2000 as International patent application Serial No. PCT/FR00/01659.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information of which I am aware and which is material to the examination of the patent application in accordance with 37 CFR §1.56.

I hereby claim foreign priority benefits under 35 U.S.C. §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate, or §365(a) of any PCT International application which designates at least one country other than the United States, listed below and have also identified below, by checking the space, any foreign application for patent or inventor's certificate, or PCT International application having a filing date before that of the application on which priority is not claimed.

Prior Foreign Application(s)

Number	Country	Day/Month/Year Filed	Priority Not Claimed
FR 99/08008	France	23/June/1999	

I hereby claim the benefit under 35 U.S.C. §119(e) of any United States provisional application(s) listed below.

Application Serial Number	Filing Date

I hereby claim the benefit under 35 U.S.C. §120 of any United States application(s), or §365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. §112, I acknowledge the duty to disclose information known to me which is material to the patentability as defined in 37 CFR §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application.

Application Serial Number	Filing Date	Status (patented, pending, abandoned)

Each undersigned applicant hereby appoints **CONRAD J. CLARK** (Registration No. 30,340) and **CHRISTOPHER W. BRODY** (Registration No. 33,613), as his attorneys with full power of substitution to prosecute the subject application and to transact all business in the Patent and Trademark Office connected therewith.

Send Correspondence to: CLARK & BRODY, 1750 K Street, NW, Suite 600, Washington, DC 20006; Telephone: 202-835-1111; Facsimile: 202-835-1755.

I hereby declare that all statements made herein of my own knowledge are true and that all statement made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or first inventor: Jamal BAINA
Inventor's signature: Date: 2 May 2002
Residence: Nancy, France
Citizenship: France
Post Office Address: 16, rue du Docteur Bernheim, F-54000 Nancy, France

Full name of second joint inventor, if any: Pierre BRETILLON
Inventor's signature: Date: 02 May 2002
Residence: Metz, France
Citizenship: France
Post Office Address: 6, rue de la Glaceire, F-57000 Metz, France

[redacted]